

2017-05-18

The Psychology of Common Knowledge Explains the Appearance of Altruistic and Moral Motivation

William H.B. McAuliffe

University of Miami, w.mcauliffe@umiami.edu

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_theses

Recommended Citation

McAuliffe, William H.B., "The Psychology of Common Knowledge Explains the Appearance of Altruistic and Moral Motivation" (2017). *Open Access Theses*. 682.

https://scholarlyrepository.miami.edu/oa_theses/682

This Embargoed is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Theses by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.

UNIVERSITY OF MIAMI

THE PSYCHOLOGY OF COMMON KNOWLEDGE EXPLAINS THE
APPEARANCE OF ALTRUISTIC AND MORAL MOTIVATION

By

William H.B. McAuliffe

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

August 2017

©2017
William H.B. McAuliffe
All Rights Reserved

UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

THE PSYCHOLOGY OF COMMON KNOWLEDGE EXPLAINS THE
APPEARANCE OF ALTRUISTIC AND MORAL MOTIVATION

William H.B. McAuliffe

Approved:

Michael E. McCullough, Ph.D.
Professor of Psychology

Elizabeth R. Losin, Ph.D.
Assistant Professor of Psychology

William A. Searcy, Ph.D.
Professor of Biology

Guillermo Prado, Ph.D.
Dean of the Graduate School

MCAULIFFE, WILLIAM H.B.

(M.S., Psychology)

The Psychology of Common Knowledge
Explains the Appearance of Altruistic and Moral Motivation

(August 2017)

Abstract of a thesis at the University of Miami.

Thesis supervised by Professor Michael E. McCullough.

No. of pages in text. (50)

Theorists of human nature have long debated whether prosocial behavior is always self-interested, or instead is at least sometimes explained by altruism or moral motivation. Experiments testing the empathy-altruism hypothesis appear to confirm the existence of altruism, while results from an experimental economic paradigm called the “dictator game” provide evidence of moral motivation. However, both experimental paradigms feature an explicit prompt to behave prosocially. Explicit prompts make the helping opportunity common knowledge among the participant, experimenter, and (sometimes) the potential recipient, and therefore confound unselfishness with a self-interested desire to avoid social censure. The present experiment ($N= 334$) recreated both the empathy-altruism and dictator game paradigms and manipulated whether the opportunity to benefit another person was explicitly prompted or merely permissible. Removing the explicit prompt dramatically reduced prosocial behavior: Giving in the dictator game paradigm disappeared completely, while giving in the empathy-altruism paradigm was attenuated and not explained by empathy. Empathy only predicted prosocial behavior when the ability to engage in prosocial behavior was common knowledge. These results undercut previous evidence that altruistic and moral considerations motivate prosocial behavior and suggest that empathy tracks the magnitude of perceived social censure from failing to help needy persons.

TABLE OF CONTENTS

	Page
LIST OF TABLES	iv
LIST OF FIGURES	v
Chapter	
1 INTRODUCTION	1
Moral Motivation: Evidence from the Dictator Game.....	2
Altruistic Motivation: Evidence from the Empathy-Altruism Hypothesis.....	3
The Psychology of Common Knowledge.....	5
Common Knowledge in the Dictator Game.....	9
Common Knowledge in the Empathy-Altruism Experiments.....	12
2 METHOD	14
Preregistration and Recruitment	14
Procedure	14
Predictions	23
3 RESULTS.....	26
Manipulation Checks	26
Analyses of Predictions	28
4 DISCUSSION	34
Implications for Altruistic and Moral Motivation	34
Limitations and Future Directions	40
Conclusion	43
References	44
Tables	48
Figures	49

LIST OF TABLES

Table 1	48
---------------	----

LIST OF FIGURES

Figure 1	49
Figure 2	50

Chapter 1: Introduction

Many influential theories of human motivation have assumed that all intentional behavior is self-interested (for a review, see Batson & Shaw, 1991). This assumption of universal egoism implies that prosocial behavior is never the product of altruistic motivation (i.e., a non-instrumental desire to benefit another person) or moral motivation (i.e., a non-instrumental desire to uphold a moral belief) (Batson, Ahmad, & Tsang, 2002).¹ Instead, all prosocial behavior is a means to an egoistic end. For example, people may behave prosocially in order to relieve their own distress, maintain a positive self-image, avoid social censure, or experience empathic joy (Batson & Shaw, 1991).

In the past few decades, however, social scientists have observed prosocial acts in experiments that do not appear to be explainable as a means to an egoistic end. The new consensus among social scientists is that altruism and moral motivation do exist, and may even be quite common (Batson & Shaw, 1991; Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007; Fehr & Schmidt, 1999). Two experimental paradigms have been especially influential in this shift of opinion: The dictator game and experiments testing the empathy-altruism hypothesis.

¹ The distinction between egoism, altruism, and moral motivation is sometimes confused with the distinction between genotypes that evolved because their phenotypic effects increased their bearers' fitness (which is often labeled "selfish") and genotypes that evolved because their phenotypic effects increased other organisms' fitness (which is often labeled "altruistic"). Although evolutionary explanations of altruistic and moral motivation are common (e.g., Baumard, André, & Sperber, 2013; Fehr & Fischbacher, 2003), each type of motive could in principle arise through either evolutionary selfishness or evolutionary altruism (Sober, 2002). Most relevant to the present investigation, researchers can study the motives underlying prosocial behavior while remaining agnostic about the evolutionary origins of the behavior.

Moral Motivation: Evidence from the Dictator Game

The most direct evidence that humans appear to care about upholding moral principles come from the results of an experimental economic game called the “dictator game” (Forsythe, Horowitz, Savin, & Sefton, 1994). The dictator game involves two participants—a “Dictator” and a “Recipient.” The Dictator is given a monetary endowment (say, \$10) and decides whether to send none, some, or all of the endowment to the Recipient; the Recipient, in contrast, has no endowment and does nothing. In the standard set-up, the game is played for one round with an anonymous stranger, features which are intended to rule out the possibility that Dictators perceive that their reputation could be affected by their decision. Because Dictators apparently cannot satisfy a desire to gain social benefits or avoid social censure by transferring money, researchers argue that non-zero transfers would instead reflect altruistic or moral motivation. And in apparent disconfirmation of the assumption that all motivation is egoistic, Dictators on average give 28% of their money away to the Recipient (Engel, 2011).

The pattern of transfers in the dictator game is more compatible with a moral motivation to uphold a principle of equity (Baumard, André, & Sperber, 2013; Camerer & Thaler, 1995; Fehr & Schmidt, 1999) than with an altruistic motivation to benefit the Recipient. For instance, the modal non-zero offer is half of the endowment (Engel, 2011), which suggests that many participants are concerned with either equity or equality (which are confounded in the standard dictator game because neither party exerts any effort to generate the endowment). Oxoby and Spraggon (2008) unconfounded considerations of equity and equality by creating treatments in which either the Dictator earned the endowment through answering test questions, or instead the Recipient answered the test

questions but the Dictator was nevertheless in charge of the dividing the money that the Recipient earned. Dictators on average gave nothing when they earned the endowment, and on average gave more than half of the endowment when the Recipient earned it. The results indicate that dictator game transfers are sensitive to equity rather than equality.

Altruistic Motivation: Evidence from the Empathy-Altruism Hypothesis

An impressive series of experiments by Daniel Batson and his colleagues have shown that when egoistic motives for helping a needy person are removed, individuals who experience less empathy tend to offer less help while those high in empathy still exhibit high levels of helping (for a review, see Batson & Shaw, 1991).² The typical experiment by Batson and his colleagues is a 2 (high empathy, low empathy) x 2 (egoistic motive, no egoistic motive) factorial design. Participants are told they will be part of an experiment that ostensibly has nothing to do with helping. For instance, one common cover story is that the participants were recruited to rate audiotapes that the college radio station is considering using for a new program. The audiotape is about an unfortunate situation that is designed to evoke empathy from participants. One commonly used story describes Katie Banks, a student whose parents suddenly died and who struggles to both attend college and support her younger siblings. Participants are told to follow certain instructions while they listen to the tape: Some are instructed to focus on the feelings and thoughts of the subject of the story (the “high-empathy” condition), whereas others are instructed to remain objective (the “low-empathy” condition). After listening to the tape, participants read a letter in which the professor in charge of the experiment presents an unexpected opportunity to volunteer time to help Katie. After reading the letter,

² Research linking empathy to altruistic motivation has characterized empathy as *feelings of sympathy or compassion in response to persons in need or distress* (Batson & Shaw, 1991).

participants indicate how many hours they are willing to help Katie on a volunteer form. The presence of an egoistic motive is manipulated by altering whether the participant can satisfy an egoistic desire by helping. For example, in one experiment (Batson, Batson, Slingsby, Harrell, Peekna, & Todd, 1991), the professor in charge of the experiment informs participants that they either will or will not learn how their help impacted Katie. (The empathic joy of learning of how the help impacted Katie is the egoistic motive in this case.)

If the empathy-altruism hypothesis is true, the level of helping should be lower in the low empathy/no egoistic motive cell than in the other three cells. This is because in the absence of empathy people will only help for egoistic reasons, but in the presence of empathy will want to help regardless of whether there are egoistic incentives. Batson identified this pattern of behavior in testing the empathy-altruism hypothesis against several rival hypothesis (for a review, see Batson & Shaw, 1991).

Several experiments show that empathy-induced helping is not (wholly) morally motivated. Wilhelm and Bekkers (2010) found that the principle of care, which states that one should help those who are in need (Hoffman, 2000), only partially mediates the relationship between empathy and helping. Van Lange (2008) had participants make several binary choice decisions about how to split a windfall between themselves and another person. Some participants experienced empathy for the other person because the other person had disclosed a sad story about themselves to the participant, while other participants did not communicate with the other person. Van Lange found that participants in the empathy condition chose better outcomes for others, but did not choose lesser outcomes for themselves (i.e., independent of the outcome for the other person) or

more equal outcomes. Participants who did not communicate with the other person tended to choose self-maximizing outcomes. It seems that empathy motivates altruism but not selflessness or egalitarianism. And in a dramatic demonstration that empathy for a victim can create altruistic motivation that conflicts with one's moral principles, Batson, Klein, Highberger, and Shaw (1995) manipulated participants to feel empathy for a particular member of a waiting list for a medical service. Despite the fact that participants in all conditions approved of the criteria by which members of the waiting list were rank-ordered, participants in the high-empathy condition took advantage of an opportunity to move the member that they experienced empathy for up on the waiting list. Insofar as the effects of empathy contradict the assumption of universal egoism, they do by providing evidence of altruistic motivation, not moral motivation.

The Psychology of Common Knowledge

Researchers routinely claim that reputational considerations cannot explain prosocial behavior in the dictator game or empathy-altruism experiments, but they may be overlooking the role that common knowledge plays in how people manage the impression they make on others. By "common knowledge" I refer to a state of affairs in which all relevant parties in a situation (1) know a certain piece of information, (2) know that all other parties know that piece of information, (3) know that all other parties know that everyone knows that piece of information, and so on *ad infinitum* (Vanderschraaf & Sillari, 2014). Compare common knowledge with private knowledge. Knowledge is merely private when one party knows a piece of information, but does not believe the information is known by others. In between private and common knowledge there are states of social knowledge that fall under the umbrella of "shared knowledge." For

instance, secondary knowledge obtains when one party knows that another party is privy to the information, but nothing more.

Note that the type of knowledge a particular party possesses is relative to what he or she knows about what the other parties know about the relevant piece of knowledge. It is possible, for example, for information to be secondary knowledge to one party, but private knowledge to a second party (e.g., perhaps the first party knows that both parties know the information, whereas the second party thinks she is uniquely privy to the information). Common knowledge is a special case in that *all* parties, not just the focal person, must know that everyone knows the piece of information, that everyone knows that everyone knows that piece of information, and so on.

Theorists and researchers have argued that common knowledge is crucial to a range of social-psychological phenomena, such as the maintenance of social norms (Hume, 1740/1976) and strategic behavior in game-theoretic dilemmas (Diekmann, 1993; Pinker, Nowak, & Lee, 2008; Schelling, 1960; Thomas, DeScioli, Haque, & Pinker, 2014). Recently, researchers have found that the likelihood that people will help others depends on whether the opportunity to help is common knowledge. For example, the bystander effect—which refers to a negative association between the number of bystanders to a situation in which someone needs help and the likelihood that any one person will offer help—may result from common knowledge among bystanders that there is an opportunity to help. Thomas, De Freitas, DeScioli, and Pinker (2016) created a virtual situation in which online participants were merchants at a market that was organized by a stall owner. All merchants could make a profit if at least one merchant forwent a fraction of his or her profit to help the stall owner upon request; if nobody

helped the stall owner, then none of the merchants could profit. The authors found that nearly all participants helped the stall owner when the request to help was private knowledge, while only about half as many participants helped when the request was common knowledge among the merchants. The authors concluded that when the request to help was private knowledge, participants helped because they believed that otherwise the stall owner would not receive help. When the request to help was common knowledge, bystanders shirked the request in the hopes that another merchant would agree to help.

Common knowledge of opportunities to help can also increase prosocial behavior. For instance, Andreoni, Rao, and Trachtman (2011) recorded the behavior of people entering a grocery store that had a Salvation Army donation solicitor at one of the store entrances. They found that a significant number of customers chose to avoid the store entrance with the solicitor if the solicitor made eye contact and politely asked for donations rather than silently ringing the Salvation Army bell without making eye contact. Those who did walk through the entrance where verbal requests were made donated more than customers who walked through the entrance where only a bell was rung. The verbal request and eye contact likely made the helping opportunity common knowledge, whereas merely walking by a ringing bell likely does not make the participant's knowledge of the helping opportunity common knowledge between the Salvation Army worker and the customer.

What bystander situations and charitable giving situations have in common is that the presence of common knowledge alters how people must respond in order to satisfy their egoistic desires. In the private knowledge condition of Thomas et al.'s (2016)

bystander situation, participants accepted paying a cost to help the stall owner because they believed that otherwise they would not profit at all. In the common knowledge condition, participants have a sufficiently high probability of profiting even if they do not pay a cost to help the stall owner. In contrast, there was little prosocial behavior in the private knowledge condition of Andreoni et al.'s (2011) field experiment because participants could keep their money without admitting to the Salvation Army worker that they did not want to donate. Participants donated more money in the common knowledge condition because none of the three options available to them could satisfy every self-interested motive they had: They could either (a) donate money (thereby sacrificing some material well-being in order to avoid social censure), (b) acknowledge and reject the request (which maximizes material well-being but incurs social censure), or (c) avoid the situation entirely (which maximizes income and avoids social censure, but sacrifices time). Participants generally preferred options (a) and (c) to (b), indicating that fear of social censure had stronger motivational force than small amounts of money.

Both the dictator game and empathy-altruism experiments make the opportunity for a sole participant to help common knowledge among the participant and experimenter, and often even the recipient as well. The presence of common knowledge in dictator games and empathy-altruism experiments would have no bearing on the thesis that dictator games and empathy-altruism experiments provide evidence of unselfish motivation if common knowledge inhibited prosocial behavior in those settings. In fact, the observation that prosocial behavior regularly occurs in both paradigms despite the inhibiting influence of common knowledge would demonstrate that altruistic and moral motives not only exist, but are also potent.

Unfortunately, common knowledge probably increases prosocial behavior in the dictator game and empathy-altruism paradigms. For if the ability to help in these paradigms was private knowledge, then self-interested participants could satisfy both the desire to maximize material well-being and the desire to avoid social censure by choosing to not behave prosocially. Only altruistically or morally motivated persons would have reason to behave prosocially, and thus observations of prosocial behavior would be strong evidence of unselfish motivation. But because the opportunity to behave prosocially in these experiments is common knowledge, selfish participants may still behave prosocially because doing so is the only way to conceal normatively inappropriate selfishness from the experimenter and recipient. Existing empirical evidence that bears out this prediction, at least with regards to the dictator game.

Common Knowledge in the Dictator Game

Researchers control for fear of social censure in the dictator game by having participants play with anonymous partners for one round only. However, Dictators may still fear social censure from the experimenter. It is also possible that Dictators transfer money to the Recipient just in case their anonymity is compromised, accidentally or otherwise. Although such a disposition may smack of paranoia, the social costs of social censure may so heavily outweigh the benefits of stinginess that a prudential motivation to avoid cheating others is beneficial in the long-run (Delton, Krasnow, Cosmides, & Tooby, 2011).

Indeed, there is empirical evidence that the level of anonymity guaranteed by the standard dictator game is not sufficient to eliminate reputational concerns. Franzen and Pointner (2012) created a modified dictator game in which neither the research assistant

nor the Recipient could know whether the Dictator's decision was the result of an intentional choice or an incidental process. Far more decisions to give little to nothing occurred than could be explained by chance, implying that many Dictators intentionally chose to keep their endowment when experimenters and potential Recipients could not tell whether the Dictator was responsible for failing to transfer money to the Recipient.

Dana, Cain, and Dawes (2006) have provided experimental evidence that common knowledge of the giving opportunity increases money transfers. They found that many potential Dictators, when given the choice, choose to receive \$9 from the experimenter instead of playing the dictator game with a \$10 endowment. The authors also found that almost no participants exited the game if they were told that the Recipient was not aware that the money that he or she may receive was from a dictator game (as opposed to a standard show-up payment). Thus, Dictators appear to be willing to pay a small cost to avoid a situation in which stinginess would generate social censure, but were not willing to pay a cost to avoid a situation in which stinginess would not incur social censure (from the Recipient). These results imply that when participants do act as Dictators (as they typically have no choice once they have agreed to participate in the study), they are transferring money to recipients in order to avoid social censure from failing to uphold the principle of equity, not to satisfy a moral motivation to comply with the principle of equity. When the opportunity to transfer money is not common knowledge, however, Dictators share less money because they do not perceive that their stinginess will cause disapproval.

It is not known whether fear of social censure is *wholly* responsible for the robust finding that Dictators on average transfer a positive amount of money to Recipients. Not

all participants in Dana et al.'s experiment (2006) chose to forgo the opportunity to play the dictator game; similarly, Franzen and Pointner's (2012) design was only able to demonstrate that transfer behavior decreases when common knowledge of the source of game's outcome is eliminated, not that all transfers are explained by reputational considerations. Both experiments are consistent with (at least) two (not mutually exclusive) possibilities: First, that there are individual differences in whether egoistic or moral motives cause positive transfers, and second, that there are individual differences in how sensitive Dictators are to the possibility that they may gain a negative reputation for keeping the entire endowment for themselves.

A field experiment by Winking and Mizer (2013) provides the clearest test to date of whether some Dictators are motivated by moral considerations. A confederate gave casino chips to pedestrians on the Las Vegas strip with the cover story of not having time to cash the chips in for money. Just before exiting the scene, the confederate either suggested that the pedestrian share the chips with another confederate sitting at a nearby bus stop bench or gave no suggestions. There was little probability of incurring social censure for not sharing the chips because the potential Recipient did not know that the participant had an opportunity to give, and the supplier of the chips did not know whether the participant acted on the opportunity to help. Consistent with the thesis that common knowledge can wholly explain positive transfers, nobody in either condition shared money with the confederate at the bus stop.

Winking and Mizer's (2013) results are strongly suggestive, but their protocol may have had features that deterred participants from acting on an intrinsic desire to share the windfall. First, participants may have viewed approaching a stranger in a

dangerous city as too risky to warrant acting on the desire to benefit the confederate. Second, even absent worries about safety, spontaneously offering to split money with a stranger is likely an awkward experience that many participants might just as soon avoid. Third, participants may not have seen the principle of equity as requiring them to share a windfall with any one stranger in particular. In contrast, the Recipient in the standard dictator game is the natural target for sharing because both parties put in the effort of attending the experiment. Consequently, it is possible that some Dictators are morally motivated to equally split monetary windfalls, but choose not to when there is a stronger egoistic reason to keep the money.

Common Knowledge in the Empathy-Altruism Experiments

Experiments testing the empathy-altruism hypothesis also make the opportunity to behave prosocially common knowledge. Specifically, participants are given an opportunity to help a needy other by an explicit prompt from the professor in charge of the experiment (e.g., Toi & Batson, 1982). While the prompt is not presented as part of the actual experiment, participants may still be motivated to appear prosocial in the eyes of the research team.

One pair of experiments in this research program (Fultz, Batson, Fortenbach, McCarthy, & Varney, 1986) tried to reduce fear of social censure by eliminating common knowledge of the link between the opportunity to help and the awareness of someone in need. In both experiments, participants were asked to form impressions of another participant, Janet, from whom they received a letter about how she has been lonely attending an out-of-state school. The participant was later given an unexpected opportunity to help by participating in another study where letter-writers and letter-

readers get to form a friendship. Participants were told that Janet will only be contacted about the second study if they chose to enroll. Janet would not know if the participant decided to not help, and the experimenter (as far as the participant could tell) did not know that Janet was in need of help. The authors found that experimentally induced empathy increased the likelihood of helping Janet, confirming the empathy-altruism hypothesis.

However, the results from Fultz et al. (1986) are not definitive for three reasons. First, the experiments had very small sample sizes.³ Second, there was still common knowledge between the experimenter and participant of an opportunity to help the experimenter by participating in the second study. Third, participants may have volunteered for the second study because they foresaw the opportunity to tell the experimenter how badly Janet needed a friend during the periodic telephone calls with the experimenter that were advertised as part of the friendship study (cf. Batson et al., 1988).

An experiment that makes the opportunity to behave prosocially private knowledge is necessary for determining whether the empathy-altruism and dictator game experiments really do provide evidence of altruistic or moral motivation. Here, I aim to test whether prosocial behavior still occurs in an experiment that recreates the core features of the dictator game and empathy-altruism paradigms, and manipulates whether the opportunity to behave prosocially is private knowledge or common knowledge.

³ Modest sample sizes are a chronic issue in the empathy-altruism literature. Many experiments had 15 or fewer participants in each cell. For example, see the experiments cited in Batson and Shaw (1991).

Chapter 2: Method

Preregistration and Recruitment

The experiment had a 2 (common knowledge vs. private knowledge) x 3 (high empathy sad letter vs. low empathy sad letter vs. neutral letter) factorial design. The sad letter conditions mimicked the circumstances of a prototypical experiment testing the empathy-altruism hypothesis, while the neutral letter conditions served as proxies of a dictator game experiment. I planned to recruit at least 50 non-suspicious participants to each of the six cells. I recruited University of Miami participants in exchange for partial course credit and the possibility of earning a \$10 bonus payment.

I ended the study with data from 377 (230 female) participants, 334 (211 female) of which believed the ruse of the experiment and were included in the final sample. Six participants that were supposed to be in the private knowledge/neutral letter condition accidentally received a sad letter. This version of the protocol did not match any of the planned conditions because the participants did not receive perspective-taking instructions. I retained these participants in the private knowledge/sad letter cell, but they were not included in the checks of the empathy manipulation. The planned sample size, study materials, predictions, and analyses were all preregistered on Open Science Framework and can be viewed here: <https://osf.io/a5wh6/>.

Procedure

Each participant waited to begin the study at a bench outside of the experiment room. The research assistant approached the participant and asked if he or she was waiting to participate in a study. When the participant replied affirmatively, the research assistant asked, “Have you seen anyone else waiting around here? The study involves two

people.” After the participant responded negatively, the research assistant said that he or she would get the participant started on the study and then look for the second person again. In reality, there was no second person, and all materials that were ostensibly authored by the second person during the study were pre-written by research assistants.

Each participant was brought to a cubicle. The participant read a consent form and decided whether to participate. The research assistant put the participant’s cell phone in a locker to minimize distractions, and then gave the participant a written introduction to the study to read while the research assistant ostensibly looked for the other participant. These instructions were presented as the introductory instructions for both the participant and the second person, so that anything written on the instructions would be common knowledge among the participant, second person, and research assistant. The research assistant informed the participant that it would take a few minutes to set up the task for the second participant, and that he or she would return after the second participant had begun the study in a different room.

The instructions informed the participant that he or she would be partnered with another person to complete tasks that would help the researchers better understand how people write and read narratives written in the first person. The instructions informed participants that each person in the pair would be randomly assigned as either the “letter-writer,” who would write an essay in the style of a memoir about recent events in his or her own personal life, or the “letter-reader,” who would provide feedback on the literary qualities of the letter-writer’s essay in order to inform further writing tasks that the letter-writer would complete. The instructions also noted that both the essay and feedback

would be sent in sealed envelopes, carried by the research assistant, to the partner who would be in a different room.

The participant then read that one member of the pair would be randomly chosen to receive \$10 as a bonus payment for participating. The instructions stated that previous experience has convinced the professor in charge of the study that offering at least \$10 is necessary for effectively recruiting participants, but that limited funding has constrained the professor to distributing money by lottery instead of giving money to all participants. In the common knowledge condition, the instructions then said, “**However**, whoever wins the lottery may feel free, if they wish, to transfer some or all of the money to their partner in the envelope with the letter (if the letter-writer wins the lottery) or the questionnaire (if the letter-reader wins the lottery).” The instructions in the private knowledge condition did not contain this proviso. Thus, in the private knowledge condition the participant had to make an inference that he or she was able to transfer the money to the other person, and the other parties did not know whether or not the participant had made that inference, nor whether the participant believed that they believed that the participant had made that inference.

Five minutes later, the research assistant returned after apparently having found and set up the second person. The research assistant put another cell phone in the locker, explaining that it belonged to the second person. The research assistant then explained that he or she had conducted the lottery for \$10 with the second person, and that the participant had won. Thus, the participant would receive \$10 and the second person would receive \$0. In the common knowledge condition (but not the private knowledge

condition), the research assistant verbally reiterated the proviso about sending money in the envelope.

Next, the research assistant conducted the “random” assignment of task roles by presenting the participant with a bowl full of folded-up slips of paper that ostensibly either said “letter-writer” or “letter-reader” (in reality, they all said “letter-reader”). The participant took a slip of paper and inevitably discovered that he or she had been assigned as letter-reader.

The research assistant asked the participant to complete two tasks while waiting for the letter-writer to complete the essay task. The first task was the Incomplete Figure Test (Torrance, 1968), which is a standard measure of creativity in which the participant draws a picture starting from a set of lines that is already on the piece of paper. The participant completed the task using colored pencils. The Incomplete Figure Test was not relevant to the present interests, but rather was introduced with the ruse that previous research had shown that creativity was related to how people read literary works.

The second task was to complete the Impression Management subscale of the Balanced Inventory of Desirable Responding (BIDR) (Paulhus, 1991) ($\alpha = .79$), a standard measure of how much individuals consciously desire to avoid social censure and gain social approval. Participant indicated how well various statements characterize them on Likert-type scales from 1 (Not at all) to 7 (Extremely well). Items include “I never cover up my mistakes” and “I have some pretty awful habits” (reverse-scored). The BIDR was also presented as measuring traits that are related to how people read literary works, but in reality was used in order to determine whether individual differences in

impression management would predict subsequent behavior. I added distractor items to the BIDR to make the fact that it measures impression management less apparent.

Paulhus (1991) advised scoring the BIDR either by summing the number of extreme scores (i.e., scores of 6 or 7), or by summing continuous scores on each item. Conceptually, the dichotomous scoring method is superior because only extreme scores are likely to reflect impression management rather than an accurate reporting of one's own traits. However, I used the continuous scoring method because Stöber, Dette, and Musch (2002) found that continuous scores yield better psychometric properties than extreme scores.

After spending a few minutes ostensibly giving the letter-writer instructions for how to complete the essay task, the research assistant returned with 10 \$1 bills and payment forms to complete. The participant was asked to complete these payment forms after completing the Incomplete Figure Test and the BIDR. Ten minutes later, the research assistant returned with a sealed envelope (ostensibly) from the letter-writer, an unused envelope, two questionnaires, and written instructions. After providing a verbal summary of what is included in the instructions, the research assistant left the participant alone.

The instructions said that the sealed envelope would contain the essay from the letter-writer, and possibly other items that the letter-writer decided to include. Each participant read that the researchers in charge of the experiment allowed the letter-writer to place anything in the envelope that will fit because they “have found in previous studies on literature and creativity that participants often like to express themselves or communicate in other ways than just the immediate task assigned to them.” The

participant was told to examine the contents of the envelope, and then to complete the questionnaires based on his or her reactions to the letter.

The first questionnaire asked the participant to rate on Likert-type scales from 1 (Not at all) to 7 (Extremely) how much he or she was feeling various emotions. Following Batson, Early, and Salvarani (1997), five adjectives on the scale (sympathetic, compassionate, moved, softhearted, and tender) constituted an index of empathy ($\alpha = .79$), and eight adjectives (alarmed, grieved, upset, worried, disturbed, perturbed, distressed, and troubled) formed a distress index ($\alpha = .92$). The word “warm” was originally intended to form part of the empathy index, but was removed because its inclusion reduced the reliability of the scale. Because empathy and distress covary but are hypothesized to generate different types of motives to help persons in need (altruistic and egoistic, respectively) (Batson et al., 1997), I controlled for distress in analyses of the effects of empathy.

The second questionnaire, which I labeled the “Feedback Questionnaire,” consisted of statements about the essay’s literary qualities. The participant was instructed to indicate his or her level of agreement with each statement on Likert-type scales from 1 (Strongly Disagree) to 7 (Strongly Agree). The ostensible purpose of the questionnaire was to provide feedback to the letter-writer on the literary qualities of the essay. However, two statements (“I thought a lot about the feelings of the narrator” and “I saw things from the perspective of the narrator”) were designed to measure how much the participant focused on the thoughts and feelings of the letter-writer while reading the essay, and two other statements (“I was pretty detached while reading this story” and “I thought about the story in an objective way”) were designed to measure the extent to

which the participant remained objective and emotionally detached while reading the letter-writer's essay. Each pair of statements was averaged and served as a manipulation check on the perspective-taking manipulation (see below). These items are not the standard items used for the manipulation check (Batson et al., 1997), but I wanted to use items that meshed well with the stated purpose of the Feedback Questionnaire.

The instructions asked participants to place the completed feedback questionnaire in the unused envelope, as well as any materials that the participant would like to send to the letter-writer that fit inside the envelope. A note or a doodle were provided as suggestions for additional materials to include (the supplies for writing notes and drawing doodles were already available because they were needed for the Incomplete Figure Test and in case the participant was assigned as the letter-writer). The participant was instructed to not reveal any identifiable information in the envelope for the ostensible purpose of standardizing the level of anonymity between the partners across study sessions. The opportunity to include additional materials was provided, of course, to give the participant an opportunity to place money in the envelope. The request to omit identifiable information was included to prevent participants from attempting to gain reputational benefits from sending money.

In the sad letter conditions, there was an additional section in the instructions titled "Reading Mindset." The instructions claim that previous research had found that the mindset that people are in when reading literature can affect their experience of it. The instructions therefore asked each participant to adopt the same mindset while reading the essay in order to control for the extraneous influence of the participant's mindset. The mindset instructions were in fact manipulated across high and low empathy groups. The

instructions we used are perspective-taking instructions from Batson et al. (1997) and have been used in several experiments to manipulate empathy.

In the high-empathy cell, participants read, “While you are reading the letter, try to imagine how the person who wrote it feels. Try to take the perspective of that person, imagining how he or she feels about what has happened and how it has affected his or her life. Try not to concern yourself with attending to all the information presented. Just concentrate on trying to imagine how the person feels.”

In the low-empathy cell, participants read, “While you are reading the letter, try to attend carefully to the information presented. Try to be as objective as possible, carefully attending to all the information presented about the situation and about the person who wrote the letter. Try not to concern yourself with how the person feels about what has happened. Just concentrate on trying to understanding objectively to the information presented.” Participants in the neutral letter condition did not receive mindset instructions.

The participant then opened the envelope, which contained the essay and a note. The note was included to reinforce the notion that the participant was indeed allowed to put materials in the envelope other than the Feedback Questionnaire. In the sad letter conditions, the essay was about a chronic financial hardship that was currently causing the letter-writer extreme distress. The attached note had a doodle of the University of Miami logo and contained an apology for the contents of the note: “Sorry for spilling my guts like that, I know this is a weird setting to do that.” There were three versions of the sad letter, and the version used in any given session was determined at random. Although

each sad letter described a serious financial issue, some notes were written to be more severe issues than others in order to enhance the generalizability of results.

In the neutral letter condition, the essay was about a leisure activity that the letter-writer desired to engage in. The attached note had the same doodle and contained an apology for possibly boring the participant: “Hope that wasn’t too boring... Looking forward to seeing your feedback!” There were two versions of the neutral letter, and the version used in any given session was determined at random.

After reading the essay and completing the questionnaires, the participant put the Feedback Questionnaire in the envelope, as well as whatever other materials he or she wished to include. Many participants (86.2%) did include notes and/or doodles in the envelope, most of which in the sad letter conditions were attempts to provide emotional social support to the letter-writer. Four research assistants blind to the purpose of the experiment rated the supportiveness of the note/doodles sent in the sad letter conditions (ICC = .84). The ratings were composed of three items, each scored on Likert-type scales from 1 (Not at all) to 7 (An extreme extent). The items asked to what extent the participant (1) expressed sympathy for the letter-writer, (2) tried to comfort the letter-writer, and (3) demonstrated genuine care for the letter-writer. An average of these ratings served as an alternative dependent variable in the sad letter conditions. I conceived of the supportiveness of the letter as representing emotional social support, or support that communicates to another person that he or she is valued, and the sending of money as representing instrumental social support, or support that provides another person with a practical means to solving his or her problem (Cohen & Wills, 1985).

The research assistant then took the envelope, with the ostensible purpose of bringing it to the letter-writer. Instead, the research assistant recorded how much money was included in the envelope. The research assistant then probed the participant for suspicion and debriefed the participant about the true nature of the experiment. The research assistant returned whatever money the participant placed in the envelope to the participant, returned the participant's cell phone, and dismissed the participant.

Predictions

My predictions were based on the hypothesis that the putative evidence of altruistic and moral motivation is, at least in part, explained by the fact that common knowledge of an opportunity to behave prosocially creates an egoistic desire to avoid social censure. Thus, my first prediction was that participants would transfer more money in the common knowledge conditions than in the private knowledge conditions.

Second, I predicted an interaction between empathy (operationalized as the contrast of the high empathy group with the low empathy group, as well as a continuous variable across the entire sample) and the common knowledge condition, such that the relationship between empathy and transferring money would be stronger in the common knowledge condition than in the private knowledge condition. Although this prediction assumes that the empathy-helping relationship is sensitive to the possibility of social censure, it is also compatible with the empathy-altruism hypothesis in that it does not rule out that empathy might still predict helping even in the absence of possible social censure.

I did not anticipate that supportive notes/doodles would become a viable alternative dependent variable, and consequently I did not generate any predictions for

how empathy would relate to supportive notes in advance of conducting the experiment. However, past research indicates that empathy increases emotional supportiveness (Fultz et al., 1986), and thus it is reasonable to predict that empathy would increase the emotional supportiveness of notes, as well as the probability of sending a note.

Third, I predicted that there would be an interaction between the common knowledge manipulation and the BIDR, such that BIDR scores would be associated with transferring money in the common knowledge condition but not in the private knowledge condition. For a participant in the common knowledge condition should perceive that he or she may incur social disapproval for not sharing money, and thus it is plausible that individual differences in sharing can be explained by individual differences in how much different participants care about appearing prosocial. Because a participant in the private knowledge condition should not perceive that he or she may incur social censure for failing to share money, trait impression management should not explain individual differences in giving behavior in the private knowledge condition.

Although I did not distinguish between the amount of prosociality (i.e., amount of money sent and the emotional supportiveness of notes/doodles) and probability of behaving prosocially at all (i.e., the decision to make a non-zero transfer and the decision to send a note/doodle, whatever its content might be) in my predictions, there is an important difference in what one may infer about the causes of variation in each type of variable. Specifically, once a person chooses to behave prosocially, the fact he or she can behave prosocially is common knowledge between the recipient and the participant, even in the private knowledge condition. Consequently, the *magnitude* of non-zero levels of prosocial behavior are always common knowledge in this context, which confounds an

egoistic desire to avoid social censure (for not being prosocial *enough*) with an unselfish (i.e., altruistic or moral) motivation to benefit the other person. In contrast, the choice to behave prosocially at all in the private knowledge condition is clear evidence of unselfish motivation because choosing to not behave prosocially does not make it common knowledge that one could have behaved prosocially.

This line of reasoning implies that the empathy-altruism hypothesis would only be supported if empathy increased the probability of behaving prosocially in both the common knowledge and private conditions. The empathy-specific punishment hypothesis, which states that empathy generates motivation to avoid social censure for failing to help needy persons (Batson et al., 1988), would be supported if empathy only predicted the magnitude of prosocial behavior, and/or if empathy only predicted the probability of sending money in the common knowledge condition.

The different incentives underlying the decision of whether to behave prosocially at all and the decision of how much prosociality to engage in also suggests that the common knowledge manipulation should increase the likelihood of making a non-zero transfer of money more than it increases the magnitude of positive transfers. Again, once someone has chosen to make a positive transfer, the ability to make a positive transfer will become common knowledge upon receipt of the transfer. So the common knowledge manipulation does not change the incentives of participants who have already decided to send money, and are just deciding how many dollars to transfer. Making the opportunity to transfer money common knowledge should affect decisions about whether to give at all because it eliminates the possibility of keeping the ability to transfer money private knowledge by not making a positive transfer.

Chapter 3: Results

I conducted all analyses using R 3.1.2 (R Core Team, 2013). All analyses used two-tailed tests. Participants were dropped only from analyses from which they had no data; in cases where participants only had partially missing data, the available data were used in the analysis. The results reported here are from participants who either (a) were not suspicious, (b) had non-specific suspicions (i.e., they believed that something about the protocol may have been inauthentic, but were not sure what aspect was inauthentic), or (c) had suspicions that were specific to the hypothesis of interest but report that their suspicions did not affect their behavior. (I refer to analyses of these participants as “valid treatment” analyses.) To examine the robustness of the results, I also conducted the analyses on all participants, suspicious or not (which I will refer to as the “intent-to-treat” analyses), as well as on only participants who were completely naïve to the experimental ruse (which I will refer to as “naïve” analyses). Except where I indicate otherwise, the results from the intent-to-treat and naïve analyses did not qualitatively differ from the valid treatment analyses.

Manipulation Checks

Did the sad letters elicit more empathy than the neutral letters?

Yes. The sad letter conditions were intended to test for the existence of empathy-induced altruism, while the neutral letter conditions were intended to test for the presence of moral motivation to uphold equity concerns. I wanted to confirm that participants indeed responded to sad letters differently than they responded to the neutral letters to confirm that the manipulation could plausibly evoke different motives. I decided to test the effect of the manipulation on empathy because the empathy-altruism hypothesis is

theorized to primarily apply to situations in which the potential beneficiary is in distress or need (Batson & Shaw, 1991). There was indeed significantly higher empathy in the sad letter condition than in the neutral letter condition, $b = 1.15$, $SE = .14$, $t(324) = 8.42$, $p < .001$, 95% CI [.88, 1.42], $d = .93$.

Did perspective-taking instructions create group differences in the mindset adopted while reading the essay?

No. Participants who were told to focus on the thoughts and feelings of the letter-writer did not report doing so to a greater extent than participants told to remain objective, $b = .13$, $SE = .15$, $t(211) = .93$, $p = .353$, 95% CI [-.15, .43], $d = .13$. Conversely, participants who were told to remain objective while reading the essay did not report doing so more than participants who were told to focus on the thoughts and feelings of the other letter-writer, $b = .20$, $SE = .19$, $t(211) = 1.05$, $p = .297$, 95% CI [-.18, .58], $d = .14$.

Did perspective-taking instructions create group differences in empathy?

No. Participants who were asked to remain objective while reading the letter-writer's essay did not report experiencing less empathy than participants who were asked to focus on the thoughts and feelings of the letter-writer while reading the essay, $b = -.18$, $SE = .17$, $t(207) = -1.05$, $p = .294$, 95% CI [-.51, .15], $d = -.15$.

Because the perspective-taking manipulation altered neither the mindset nor the reported empathy of participants, I was not able to test the causal effect of empathy on money transfers or the supportiveness of notes. Although the sad letter manipulation did causally increase empathy, manipulating the perceived severity of the beneficiary's need

also increases egoistic motives in tandem (Batson, 1997). Thus, the sad letter manipulation cannot serve as a manipulation of empathy in this design.

I decided to test my predictions about the effects of empathy using self-reported empathy as a continuous variable, which can at least test whether the results are consistent with the empathy-altruism hypothesis. I collapsed the high-empathy and low-empathy groups together to form one sad letter condition. Thus, the analyses reported below are effectively from a 2 (sad letter vs. neutral letter) X 2 (common knowledge vs. private knowledge) factorial design. In all cases where the common knowledge manipulation was used as an independent variable, the common knowledge condition was dummy coded as 1, and the private knowledge condition was dummy coded as 0. In all cases where the sad letter manipulation was used as an independent variable, the sad letter condition was dummy coded as 1, and the neutral letter condition was dummy coded as 0.

Analyses of Predictions

The confirmatory analyses all had a zero-inflated count variable—money transferred in the envelope—as the dependent variable (see Figures 1 and 2 for indications that the distribution was zero-inflated). I realized that models that were designed for skewed count data might fit the data better than the standard ordinary least-squares regression model (Atkins, & Gallop, 2007). Using the Akaike information criterion (AIC) as my measure of model fit, I fit an ordinary least squares model (AIC = 1845.88), Poisson model (AIC = 2264.5), negative binomial model (AIC = 1427.99), zero-inflated Poisson model (AIC = 1163.58), and zero-inflated negative binomial model (AIC = 1164.43) to a regression analysis in which the common knowledge manipulation

predicted the amount of money transferred (I chose these variables since they are both present in almost all of the confirmatory analyses).

The zero-inflated Poisson model had the lowest AIC, and was accordingly used in the analyses in which money was the dependent variable. Note that zero-inflated models produce both logistic regression coefficients and count regression coefficients. Logistic coefficients indicate whether the log-odds of observing a 0 (in this context, this means no money was transferred) is higher than the odds of observing a 1 (in this context, this means that at least \$1 was transferred). Below I report the exponentiated coefficient, the odds ratio (OR), in addition to the raw coefficient to facilitate ease of interpretation. (Standard binary logistic regression predicts the odds of observing 1 rather than 0, and accordingly the signs of coefficients are reversed in zero-inflated Poisson regression.) Count coefficients indicate the increase in log-rate of the outcome (in this context, an increase in amount of money transferred) for each one-unit increase in the predictor among the non-zero data (in this context, among participants who transferred at least \$1). Below I report the exponentiated coefficient, the rates ratio (RR), in addition to the raw coefficient to facilitate ease of interpretation. Logistic coefficients typically reflect different psychological processes than count coefficients (Atkins, & Gallop, 2007). In the present context, significant, negative logistic coefficients may provide evidence of unselfish motivation (at least in the private knowledge condition), whereas significant, positive count coefficients most plausibly provide evidence of fear of social censure.

Did participants in the common knowledge condition transfer more money?

Yes. (See Table 1, Figure 1, and Figure 2.) I created a multiple linear regression model of the effects of the common knowledge manipulation and the sad letter

manipulation on amount of money transferred. Examining the logistic coefficients, the intercept was significant, $b = 1.66$, $SE = .28$, $Z(331) = 6.05$, $OR = 5.28$, 95% CI [3.08, 9.05], $p < .001$, indicating that the odds of transferring no money was over 5 times greater than the odds of making a non-zero transfer in the neutral letter/private knowledge cell. The effect of the sad letter was significant, $b = -.75$, $SE = .27$, $Z(331) = -2.80$, $OR = .47$, 95% CI [.28, .80], $p = .005$, indicating that the odds of making a non-zero transfer in the sad letter condition was about twice as high as in the neutral letter condition. The effect of common knowledge was significant, $b = -2.01$, $SE = .25$, $Z(331) = -7.92$, $OR = .13$, 95% CI [.08, .21], $p < .001$, suggesting that the odds of making a non-zero transfer in the common knowledge condition was over seven times higher than in the private knowledge condition.

Examining the count coefficients, the intercept was significant, $b = 1.68$, $SE = .08$, $Z(331) = 19.53$, $RR = 5.41$, 95% CI [4.57, 6.41], $p < .001$, indicating that there is a 441% increase in the private knowledge/neutral letter condition in giving compared to participants who did not send money. The sad letter manipulation was also significant, $b = .44$, $SE = .08$, $Z(331) = 5.76$, $RR = 1.55$, 95% CI [1.34, 1.80], $p < .001$, indicating that there was a 55% increase in giving in the sad letter condition compared to the neutral letter condition. The common knowledge manipulation was not significant, $b = -.08$, $SE = .07$, $Z(331) = -1.24$, $RR = .92$, 95% CI [.80, 1.04], $p = .214$, indicating that the magnitude of positive transfers in the common knowledge condition was not greater than transfers in the private knowledge condition.

Next, I added the interaction term of the sad letter manipulation and the common knowledge manipulation to the model. The interaction term in the logistic model was not

significant, $b = .12$, $SE = .56$, $Z(330) = .21$, $OR = 1.12$, $95\% CI [.37, 3.35]$, $p = .830$, indicating that the effect that the positive effect that the common knowledge manipulation has on making a non-zero transfer did not differ as a function of whether or not the participant read a sad letter or a neutral letter. The interaction term in the count model was also nonsignificant, $b = -.12$, $SE = .19$, $Z(330) = -.64$, $RR = .88$, $95\% CI [.61, 1.29]$, $p = .520$, indicating that the common knowledge manipulation does not affect the size of positive transfers in either the sad letter or neutral condition.

Did empathy predict money transfers more strongly in the common knowledge condition than in the private knowledge condition?

No, empathy only predicted the magnitude of money transferred. I first examined the main effects of empathy, distress, and the common knowledge manipulation on money transferred. Examining the logistic model coefficients, I found that neither empathy, $b = -.16$, $SE = .11$, $Z(322) = -1.40$, $OR = .85$, $95\% CI [.70, 1.28]$, $p = .163$, nor distress, $b = -.10$, $SE = .08$, $Z(322) = -1.27$, $OR = .90$, $95\% CI [.78, 1.05]$, $p = .204$, significantly predicted the odds of making a non-zero transfer. (However, empathy was a significant predictor in the intent-to-treat analysis: $b = -.20$, $SE = .10$, $Z(366) = -2.00$, $OR = .81$, $95\% CI [.66, .99]$, $p = .046$.) Examining the count model coefficients, I found that empathy, $b = .10$, $SE = .03$, $Z(322) = 3.53$, $RR = 1.11$, $95\% CI [1.05, 1.18]$, $p = <.001$, and distress, $b = .06$, $SE = .02$, $Z(322) = 3.02$, $RR = 1.06$, $95\% CI [1.02, 1.09]$, $p = .002$, were both positively associated with the rate of money transfers.

Next, I reran the model with an added interaction between the common knowledge manipulation and empathy. Examining the logistic model, the interaction of empathy and the common knowledge manipulation was not significant, $b = -.19$, $SE = .20$,

$Z(321) = -.98$, $OR = .82$, $95\% CI [.56, 1.22]$, $p = .329$. Similarly, the count model revealed that the interaction of empathy and the common knowledge manipulation was not significant, $b = .06$, $SE = .06$, $Z(321) = 1.09$, $RR = 1.07$, $95\% CI [.95, 1.19]$, $p = .277$.

Did empathy predict the supportiveness of notes?

Yes. I ran a multiple regression model with empathy and distress predicting the supportiveness of notes sent in the sad letter condition (recall that raters only examined notes/doodles sent in the sad letter condition). The overall model was significant, $F(2, 178) = 10.69$, $p < .001$, $R^2_{adj} = .10$. There was a significant effect of empathy, $b = .14$, $SE = .05$, $t(178) = 2.97$, $p = .003$, $95\% CI [.05, .25]$, $d = .44$, as well as a significant effect of distress, $b = .06$, $SE = .03$, $t(178) = 2.17$, $p = .031$, $95\% CI [.01, .13]$, $d = .32$.

Next, I created a binary logistic regression model with empathy and distress predicting the odds of sending a note in all conditions. Neither the effect of empathy, $b = .19$, $SE = .14$, $Z(323) = 1.35$, $OR = 1.21$, $95\% CI [.91, 1.58]$, $p = .178$, nor distress, $b = .00$, $SE = .20$, $Z(323) = -.04$, $OR = 1$, $95\% CI [.82, 1.22]$, $p = .972$, was significant. (However, empathy did significantly increase the odds of sending a note/doodle in the intent-to-treat analysis, $b = .34$, $SE = .13$, $Z(366) = 2.65$, $OR = 1.40$, $95\% CI [1.09, 1.82]$, $p = .008$.) We reran the analysis with an added interaction between the empathy and the sad letter manipulation, but the interaction was not significant, $b = .35$, $SE = .32$, $Z(321) = 1.09$, $OR = 1.21$, $95\% CI [.76, 2.65]$, $p = .274$.

Notably, a binary logistic regression revealed that participants in the sad letter condition were nonsignificantly *less* likely to send a note/doodle than participants in the neutral letter condition, $b = -.32$, $SE = .35$, $Z(332) = -.90$, $OR = .73$, $95\% CI [.31, 1.49]$, $p = .367$. This finding is in contrast to the finding that there were, on average, positive

money transfers in the sad letter/private knowledge cell. I wondered whether the discrepancy reflected a difference in motives for sending notes/doodles versus sending money. Consistent with this possibility, I found that the supportiveness of notes did not predict sending money, $b = .07$, $SE = .15$, $Z(178) = .48$, $OR = 1.07$, $95\% CI [.80, 1.43]$, $p = .633$, or the magnitude of non-zero transfers, $b = .05$, $SE = .04$, $Z(178) = 1.46$, $RR = 1.05$, $95\% CI [.98, 1.13]$, $p = .144$. Similarly, whether or not participants sent a note did not increase the probability of sending money, $b = -.24$, $SE = .13$, $Z(323) = -.75$, $OR = .79$, $95\% CI [.41, 1.47]$, $p = .454$, or the magnitude of non-zero transfers, $b = -.09$, $SE = .09$, $Z(323) = -.97$, $RR = .91$, $95\% CI [.77, 1.09]$, $p = .332$.

Did the BIDR predict money transfers in the common knowledge condition?

No. First, I created a multiple regression model of the effects of the common knowledge manipulation and the BIDR on the amount of money transferred. The logistic model revealed that the effect of the BIDR was nonsignificant, $b = .00$, $SE = .01$, $Z(330) = .01$, $OR = 1.00$, $95\% CI [.99, 1.01]$, $p = .993$, as did the count model, $b = .00$, $SE = .01$, $Z(330) = .343$, $RR = 1.00$, $95\% CI [.99, 1.00]$, $p = .732$.

Next, I reran the model with an added interaction term of the common knowledge manipulation and the BIDR. The logistic model revealed that the interaction of BIDR and the common knowledge manipulation was nonsignificant, $b = -.00$, $SE = .02$, $Z(320) = -.01$, $OR = 1.00$, $95\% CI [.97, 1.03]$, $p = .993$, as did the count model, $b = .01$, $SE = .01$, $Z(320) = .32$, $RR = 1.00$, $95\% CI [.99, 1.01]$, $p = .752$.

Chapter 4: Discussion

Analyses from the present experiment yielded four primary findings. First, the common knowledge manipulation increased the likelihood of making a positive money transfer (but not the magnitude of positive transfers). When the giving opportunity was common knowledge, participants gave in both the sad letter and neutral letter conditions. In contrast, when the giving opportunity was private, giving was attenuated in the sad letter condition and not significantly different from zero in the neutral letter condition. Second, empathy predicted the magnitude of money transfers, but not the likelihood of making a positive transfer. Third, empathy predict the supportiveness of notes and doodles sent in the sad letter condition, but did not the probability of sending a note/doodle in any condition. Fourth, the BIDR did not predict individual differences in money transfers.

Implications for Altruistic and Moral Motivation

What explains dictator game transfers?

When participants read a neutral letter in the common knowledge condition, the pattern of money transfers looked very similar to a standard dictator game experiment (Engel, 2011): Participants (a) on average transferred 30% of their endowment, (b) had a modal transfer of zero dollars, (c) and had a non-zero modal transfer of half of the endowment (see Table 1 and Figure 1). This pattern of findings strengthens my confidence that I had recreated the essential features of the standard dictator game, allowing me to test whether canonical dictator game findings are the product of the common knowledge of the opportunity to share money.

In the private knowledge/neutral letter cell, participants on average did not transfer any money to the letter-writer. This finding suggests that the potential for social censure explains non-zero transfers in the common knowledge condition. Thus, the present results comport with other experiments which suggest that the dictator game measures a desire to *appear* equitable, not an intrinsic desire to *be* equitable (Bardsley, 2008; Dana et al., 2006; Franzen & Pointner, 2012; List, 2007; Winking & Mizer, 2013). In this face of these data, theories that draw on the dictator game as evidence that humans are morally motivated to achieve equitable outcomes (e.g., Baumard et al., 2013; Fehr & Schmidt, 1999) are either false or require revision.

Is it possible that participants in the neutral letter/private knowledge cell simply did not know that they *could* put money in the envelope? After all, participants in the private knowledge condition were not told that they could put money in the envelope, but had to infer it from the fact they could put in anything that could fit inside of the envelope. However, there are at least two reasons for disfavoring this explanation of non-significant money transfers in the neutral letter/private knowledge. First, participants in the sad letter/private knowledge cell did on average make non-zero transfers, despite not being given any more information about what they could put in the envelope than participants in the neutral letter/private knowledge cell. Second, out of the total 13 participants in the valid treatment analyses that professed ignorance to the ability to put money in the envelope (which occurred during the post-experimental interview, after learning about the true nature of the experiment), only three of these participants were in the neutral letter/private knowledge cell.

Eight participants in the private knowledge condition *did* transfer five dollars to the letter-writer. This finding confirms Winking and Mizer's (2013) statistical conclusion that Dictators on average do not transfer money to the Recipient, while failing to replicate their observation of zero sharing whatsoever. Participants in the neutral letter/private knowledge condition who did transfer money may have been morally motivated, or else especially sensitive to the possibility of receiving social censure for failing to share money. Whatever motivated these eight participants to give, it appears that Winking and Mizer's (2013) design may have discouraged a small minority of participants from expressing a desire to share their endowment.

When does empathy predict prosocial behavior?

The empathy-altruism hypothesis predicts that empathy will cause prosocial behavior both in the presence and absence of an egoistic incentive (Batson & Shaw, 1991). In contrast, the empathy-specific punishment hypothesis predicts that empathy will predict prosocial behavior when it is possible to incur social censure for failing to behave prosocially, but not otherwise (Batson et al., 1988). Because empathy was not successfully manipulated, I was not in a position to make causal claims about the effects of empathy on prosocial behavior. However, I was able to test whether the results are at least consistent with the empathy-altruism hypothesis. Regardless of whether money transfers or notes/doodles were used as the dependent variable, the empathy-specific punishment hypothesis received support, and the empathy-altruism hypothesis was disconfirmed: Empathy predicted the magnitude of positive money transfers, but not the probability of making a positive transfer. This finding suggests that when the opportunity to give was common knowledge, the experience of empathy motivated larger transfers of

money to the letter-writer, perhaps by drawing attention to the social censure that the participant might incur for failing to provide sufficient help someone in great financial need. When the opportunity to help was private knowledge, there was a much smaller likelihood of incurring social censure for failing to help, and thus empathy did not motivate sending money.

The relationship between empathy and emotional social support painted a similar picture. If the empathy-altruism hypothesis is true, then empathy should have predicted the probability of sending a note/doodle because the ability to do so was not common knowledge between the participant and the letter-writer. Although both the letter-writer and participant were given the ability to send notes and doodles, they were not able to infer that their partner knew as much because they received role-specific instructions for sending envelopes to each other (and therefore did not see what each other's instructions said regarding what is permissible to put inside the envelope). However, empathy only predicted the supportiveness of notes/doodles that were sent, not the probability of sending a note/doodle. This pattern of results is consistent with the empathy-specific punishment hypothesis because the opportunity to send a doodle/note was common knowledge once the participant decided to send a note/doodle.

An interesting caveat to the above reasoning is that empathy *did* significantly increase the odds of both sending a note/doodle and of making a non-zero monetary transfer in the intent-to-treat analyses. One might want to privilege the intent-to-treat analyses insofar as they have more statistical power and do not violate random assignment. On the other hand, the suspicious participants had a fundamentally different experience than was intended. One would expect that there would be little association

between emotions and prosocial behavior among suspicious participants because suspicious participants are less immersed in the experimental situation. The fact that suspicious participants' knowledge of the true purpose of the experiment *increased* the observed association between empathy and the odds of behaving prosocially is very surprising.

However, the empathy-specific punishment hypothesis can easily explain this empirical anomaly. Many participants who were suspicious of the ruse believed that they were being evaluated according to how prosocially they behaved. The empathy-specific punishment hypothesis predicts that participants who are acutely aware that they are being evaluated should experience an empathy-induced desire to behave prosocially in order to avoid social evaluation. The fact that including suspicious participants makes the effect of empathy on the probability of behaving prosocially statistically significant is therefore consistent with the empathy-specific punishment hypothesis.

If the empathy-specific punishment hypothesis is possibly true, then why did researchers dismiss it in the 1980s? One possibility is that experiments that purported to disconfirm the empathy-specific punishment hypothesis (Batson et al., 1988; Fultz et al., 1986) were underpowered due to very small sample sizes. Another possibility is that the empathy-specific punishment hypothesis *has* received empirical support, but has not been labeled as such. For instance, one pair of experiments concluded that valuing the person in need is a prerequisite to empathy-induced altruism (Batson, Eklund, Chermok, Hoyt, & Ortiz, 2007). However, the so-called valuation manipulation was information about the needy person's moral character. Because people tend to "typecast" others as either only perpetrators or only victims (Gray & Wegner, 2009), participants likely believed that they

would not incur censure for failing to help someone of low moral character. If so, then the empathy-specific punishment hypothesis would predict that the empathy-helping relationship would only hold toward deserving targets.

Why were there money transfers in the sad letter/private knowledge condition?

The pattern of money transfers did not present an entirely hopeless picture for the skeptic of universal egoism. There was a statistically significant amount of money transferred to the letter-writer in the sad letter/private knowledge cell, suggesting that common knowledge of the giving opportunity is not a necessary condition for prosocial behavior. There is no clear answer to what motivated these transfers. One possibility is that participants inferred that the interaction partner was subtly asking them to transfer money by mentioning a financial issue. If so, then perhaps participants in this cell did experience some social pressure to transfer money. On the other hand, the ability to give even in this case would not be common knowledge. Another plausible egoistic explanation is that some participants desired to avoid guilt (Estrada-Hollenbeck & Heatherton, 1998). Indeed, Batson (1997) argues that manipulations of a beneficiary's need level also alter the anticipated social or self-censure for failing to help. The difference in anticipated guilt may explain why there was non-zero giving on average in the sad letter/private knowledge cell and zero giving on average in the neutral letter/private knowledge cell.

Moving past skeptical arguments against unselfishness, it is possible that giving in this cell was motivated by altruism. However, empathy did not increase the probability of positive transfers, and yet empathy is the only known (potential) elicitor of altruistic motivation. Another possibility is that the transfers may resulted from a moral motivation

to help those in need (Hoffman, 2000), which would explain why there was only significant giving in the sad letter condition. Future research is required to disentangle what motivates spontaneous helping of persons in need.

Why did the BIDR fail to explain individual differences in money transfers?

I predicted that impression management would explain individual differences in money transfers in the common knowledge condition. I found that the BIDR did not predict the probability the positive money transfers or the magnitude of money transfers in either level of the common knowledge manipulation. I conclude that the strength of the common knowledge manipulation did not depend on the trait level of participants' tendency to provide socially desirable responses.

This null finding was surprising because the common knowledge manipulation presumably increases prosocial behavior by increasing the perceived social censure from failing to behaving prosocially. Perhaps differences in the strength of *belief* that the transfer decision would be socially evaluated explains individual differences in money transfers. If true, then maybe empathy, which did predict the magnitude of money transfers, varies according to one's confidence that the decision to behave prosocially will be evaluated by others. An important caveat is that high scores on the BIDR reflect deliberate lying, whereas behaving prosocially in the present context reflects an honest commitment to benefitting the other person. Even if the BIDR and sending money both reflect impression management, those who behave prosocially to maintain social status may not be willing to lie in order to save face.

Limitations and Future Directions

There are (at least) four plausible threats to the generalizability of the present findings that should be addressed in future studies. First, the design was only intended to measure moral motivation to uphold a principle of equity, not any moral belief whatsoever. I focused on motivation to uphold equity principles because it has received much theoretical attention (Baumard et al., 2013; Fehr & Schmidt, 1999), and because participants (appear to) care about equitable outcomes when playing the dictator game. However, it is possible that humans do possess moral motivations to uphold other moral beliefs, such as the principle of care, even if they are not motivated to uphold their beliefs about equity.

Second, it is possible that humans do not have altruistic or moral motives to benefit strangers, but do have non-egoistic motives to benefit family members or friends. There are at least two plausible explanations for why researchers have focused on investigating the motives underlying prosocial behavior among strangers rather than among persons who are well-acquainted. First, it is easier to rule out certain kinds of egoistic explanations for prosocial behavior using strangers instead of relatives or friends. For example, it is difficult to rule out the possibility that a prosocial act toward a family member is explained by considerations of reciprocity, given that relationships between family members typically extend indefinitely into the future. Second, researchers have likely assumed that any evidence of moral or altruistic motivation toward strangers would extend, *a fortiori*, toward family members and friends, whereas the converse does not hold.

Third, it is possible that the motives that underlie prosocial behavior differ across cultures. However, as pointed out by Gaertner, Sedikides, Cai, and Brown (2010), many behavioral differences among cultures are the product of the same psychological processes. For instance, despite the enormous amount of variation that Henrich et al. (2010) observe in dictator game transfers, they conclude that the cause of this variability appears to relate to a difference in beliefs about which sharing norm applies to the dictator game, not to differences in the type of motivation that leads to adhering to whatever the relevant norm happens to be. Thus, I would predict that the motives observed in the present experiment would replicate across cultures, even if the behavioral expression of those motives differs when people from different cultures are placed in the same situation. Of course, only future empirical research can definitively resolve this matter.

Fourth, most experiments testing the empathy-altruism hypothesis present the opportunity to help *after* presenting the needy person and measuring emotional responses to the needy persons (e.g., Batson et al., 2007). The logic behind this ordering of events is to prevent the knowledge of an upcoming opportunity to help from influencing how participants react to learning about the person in need (as well as how truthfully they report those reactions). In contrast, participants in the present experiment knew how to benefit the other person before learning about their needs and before reporting emotional reactions to learning about those needs. It is possible that participants downregulated their emotional reactions to the letter-writer so that they would not feel compelled to help him or her (Cameron & Payne, 2011). If so, empathy may have had a weaker effect on helping behavior in the present experiment than if the helping opportunity had been

unexpected. Future research is required to determine whether the temporal ordering of the presentation of the needy person, measurement of emotional reactions, and presentation of helping opportunity has a causal influence on the empathy-helping relationship.

Conclusion

Results from the dictator game and empathy-altruism experiments have caused a sea change in researchers' perceptions of the desires that underlie prosocial behavior. However, the present results challenge the current consensus that altruistic motivation and a moral motivation to uphold the principle of equity explain human prosocial behavior. Common knowledge of the opportunity to behave prosocially—which can induce an egoistic desire to avoid social censure for failing to behave prosocially—may explain empathy-induced helping and equitable sharing. People *do* apparently sometimes desire to give to those in need, even when the opportunity to do so is not common knowledge, but because giving in this situation is not explained by empathy the motives that underlie such instances of prosocial behavior remain unknown.

References

- Andreoni, J., Rao, J. M., & Trachtman, H. (2011). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving (No. w17648). National Bureau of Economic Research.
- Atkins, D. C., & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: a tutorial on count regression and zero-inflated models. *Journal of Family Psychology*, 21(4), 726-735.
- Bardsley, N. (2008). Dictator game giving: altruism or artefact?. *Experimental Economics*, 11(2), 122-133.
- Batson, C. D. (1997). Self–other merging and the empathy–altruism hypothesis: Reply to Neuberg et al. (1997). *Journal of Personality and Social Psychology*, 73(3), 517-522.
- Batson, C. D., Ahmad, N., & Tsang, J. A. (2002). Four motives for community involvement. *Journal of Social Issues*, 58(3), 429-445.
- Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, 61(3), 413-426.
- Batson, C. D., Dyck, J. L., Brandt, J. R., Batson, J. G., Powell, A. L., McMaster, M. R., & Griffitt, C. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, 55(1), 52-77.
- Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus imagining how you would feel. *Personality and Social Psychology Bulletin*, 23, 751–758.
- Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., & Ortiz, B. G. (2007). An additional antecedent of empathic concern: valuing the welfare of the person in need. *Journal of personality and social psychology*, 93(1), 65-74.
- Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L.L. (1995). Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology*, 68, 1042-1054.
- Batson, C. D., & Shaw, L. L. (1991). Evidence for altruism: Toward a pluralism of prosocial motives. *Psychological Inquiry*, 2(2), 107-122.

- Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36(01), 59-78.
- Camerer, C., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, 9(2), 209-219.
- Cameron, C. D., & Payne, B. K. (2011). Escaping affect: how motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, 100(1), 1-15.
- Cohen, S., & Wills, T. A. (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin*, 98(2), 310-357.
- Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior And Human Decision Processes*, 100(2), 193-201.
- Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences*, 108(32), 13335-13340.
- Diekmann, A. (1993). Cooperation in an asymmetric volunteer's dilemma game: Theory and experimental evidence. *International Journal of Game Theory*, 22, 75– 85.
- Engel, C. (2011). Dictator games: A meta-study. *Experimental Economics*, 14(4), 583-610.
- Estrada-Hollenbeck, M., & Heatherton, T. F. (1998). Avoiding and alleviating guilt through prosocial behavior. In J. Bybee (Ed.), *Guilt and children* (pp. 215– 231). San Diego, CA: Academic Press.
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425, 785-791.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817-868.
- Forsythe, R., Horowitz, J., Savin, N., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3), 347-369.
- Franzen, A., & Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior & Organization*, 81(1), 74-81.

- Fultz, J., Batson, C. D., Fortenbach, V. A., McCarthy, P. M., & Varney, L. L. (1986). Social evaluation and the empathy–altruism hypothesis. *Journal of Personality and Social Psychology*, 50(4), 761-769.
- Gaertner, L., Sedikides, C., Cai, H., & Brown, J. D. (2010). It's not WEIRD, it's WRONG: When Researchers Overlook uNderlying Genotypes, they will not detect universal processes. *Behavioral and Brain Sciences*, 33(2-3), 93-94.
- Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, 96(3), 505-520.
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., & Lesorogol, C. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972), 1480-1484.
- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge, UK: Cambridge University Press.
- Hume, D. (1740/1976). *A treatise of human nature*. Oxford, UK: Clarendon Press.
- List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3), 482-493.
- Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, 65(3-4), 703-713.
- Paulhus, D.L. (1991). Measurement and control of response biases. In J.P. Robinson et al. (Eds.), *Measures of personality and social psychological attitudes*. San Diego: Academic Press.
- Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of sciences*, 105(3), 833-838.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: <http://www.R-project.org/>.
- Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press: Cambridge, MA.
- Sober, E. (2002). The ABC of altruism. In S. G. Post, L. G. Underwood, J. Schloss, & P. W. B. Hurlbut (Eds.), *Altruism and altruistic love* (pp. 17–28). Oxford, England: Oxford University Press.

- Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment*, 78(2), 370-389.
- Thomas, K. A., De Freitas, J., DeScioli, P., & Pinker, S. (2016). Recursive mentalizing and common knowledge in the bystander effect. *Journal of Experimental Psychology: General*, 145(5), 621-629.
- Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, 107(4), 657-676.
- Toi, M., & Batson, C. D. (1982). More evidence that empathy is a source of altruistic motivation. *Journal of Personality and Social Psychology*, 43(2), 281-292.
- Torrance, E. P. (1968). *Torrance tests of creative thinking*. Personnel Press, Incorporated.
- Van Lange, P. A. (2008). Does empathy trigger only altruistic motivation? How about selflessness or justice? *Emotion*, 8(6), 766-774.
- Vanderschraaf, P., & Sillari, G. (2014). Common knowledge. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2014/entries/common-knowledge/>.
- Wilhelm, M. O., & Bekkers, R. (2010). Helping behavior, dispositional empathic concern, and the principle of care. *Social Psychology Quarterly*, 73(1), 11-32.
- Winking, J., & Mizer, N. (2013). Natural-field dictator game shows no altruistic giving. *Evolution and Human Behavior*, 34(4), 288-293.

Table 1. Money transfers in each condition

Cell	<i>N</i>	Mean	<i>SD</i>	Max	Mode
Common knowledge/Sad letter	103	5.75	4.04	11	10
Private knowledge/Sad letter	117	2.47	4.13	15	0
Common knowledge/Neutral letter	61	3.02	3.21	10	0
Private knowledge/Neutral letter	53	0.75	1.81	5	0

Note: Units are unstandardized (in dollars). The maximum transfer in the sad letter conditions was higher than \$10 due to two participants who, in addition to putting the \$10 they received from the experiment in the envelope, also included money that they had on their person. Including these cases in the data analysis did not qualitatively change the results (as compared to changing the values to \$10.)

Figure 1. Money transfers at each level of common knowledge manipulation in the neutral letter condition

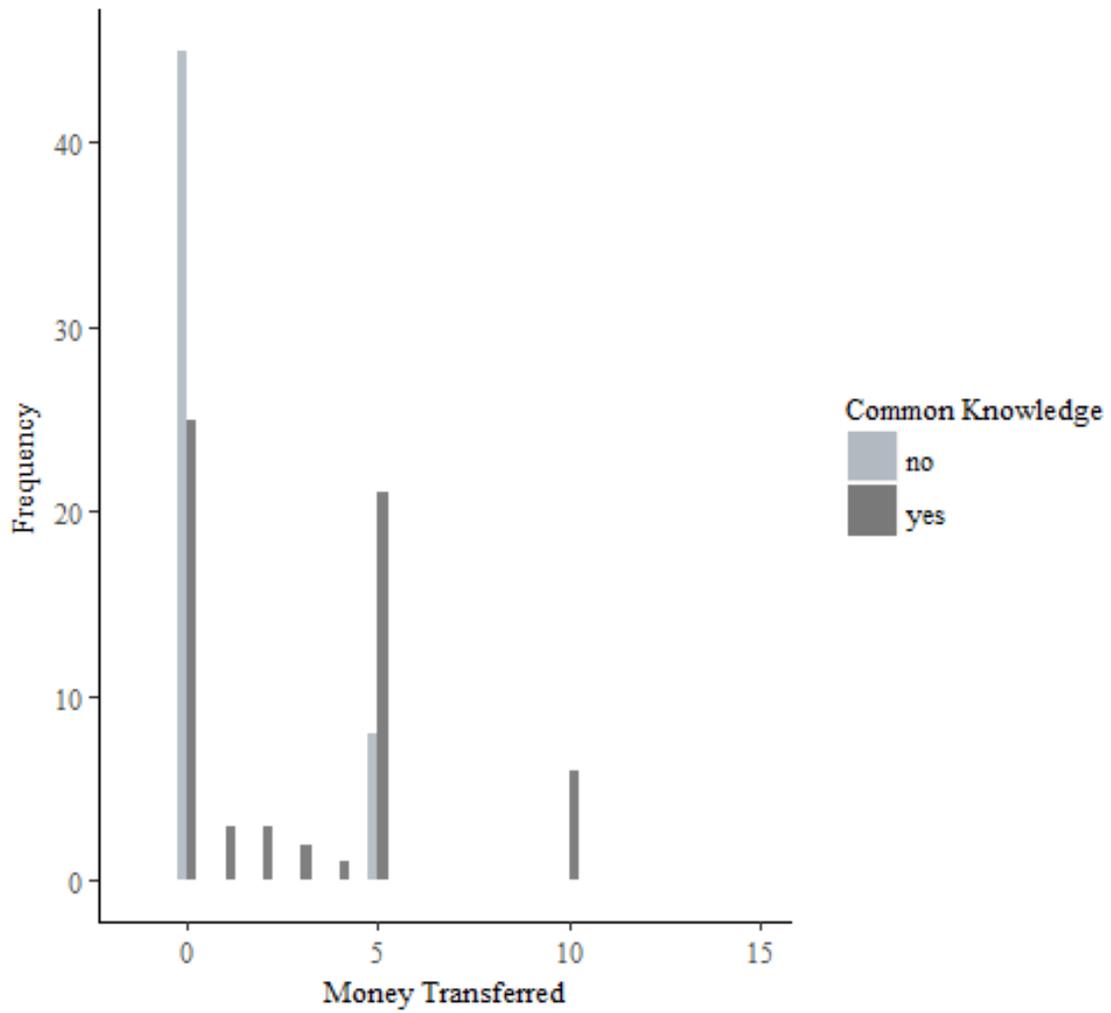


Figure 2. Money transfers at each level of common knowledge manipulation in the sad letter condition

